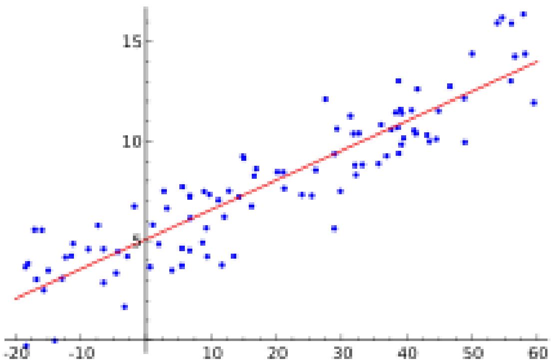
广义线性模型 学习笔记（一）——定义

0.引言

当对数据进行回归分析时，经常会用到两类回归模型：

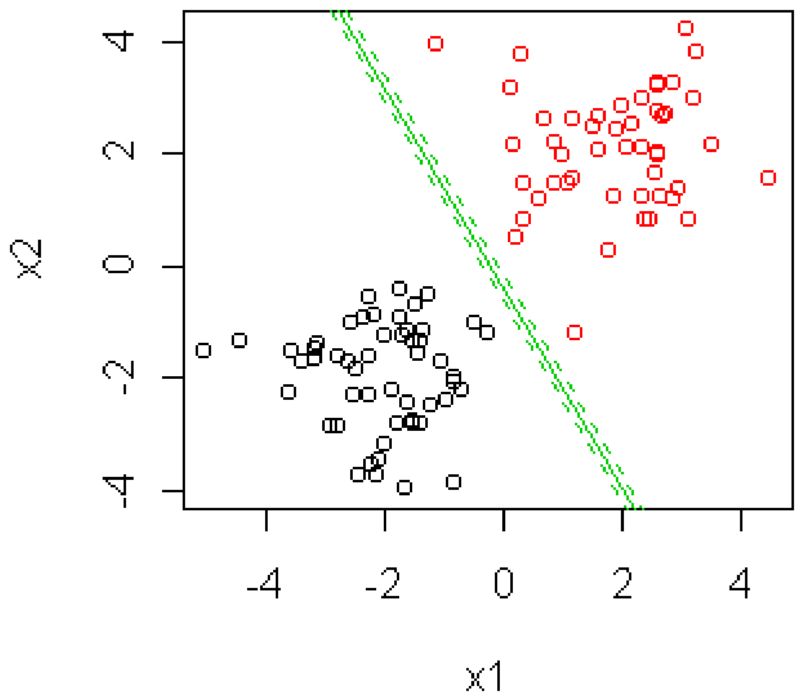
一类是针对连续型数据进行的线性回归，如下图（图片来源：维基百科）所示



这种回归模型的公式是：

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

还有一类是针对分类数据进行的logistic回归，如下图（图片来源）



这种回归模型的公式则是：

$$y = \frac{1}{1 + e^{-(a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n)}}$$

这里形如

$$y = \frac{1}{1 + e^{-z}}$$

的函数被称为 sigmod 函数。

那么问题来了：这两种回归模型，是否有可能统一到一种更加抽象的模型中？这两种形式完全不一样的模型，是不是只是某种更加广义的模型的特例？

答案是肯定的，这个更加广义、抽象的模型就叫做广义线性模型(Generalized Linear Model)，简称 GLM。 常见的线性回归、logistic 回归、泊松回归等等都是它的特殊形式。

不仅如此，在 GLM 的视角下，我们能够合理地解释：为什么logistic 回归要用sigmod 函数 $y = \frac{1}{1+e^{-z}}$？为什么泊松回归要用指数函数$y=e^{z}$？选择这些函数不仅仅是因为他们的函数曲线形状与观测值很接近，其背后有更加深层次的统计基础。

1. 广义线性模型的定义

广义线性模型建立在三个定义的基础上，分别为：

- 定义线性预测算子

$$\eta = \theta^T x$$

- 定义y 的估计值

$$h(x, \theta) = E(y|x, \theta)$$

- 定义 y 的估值概率分布属于某种指数分布族：

$$\Pr(y|x, \theta) = b(y) exp(\eta^T T(y) - a(\eta))$$

接下来详细解释各个定义

1.1. 定义一：线性预测算子

广义线性模型的名字中有『线性』两个字，自然它包含了线性计算的过程，也就是它的假设之一，定义线性预测算子(linear predictor)为：

$$\eta = \theta^T x$$

这里的 θ 和x都是向量，写成标量形式就是：

$$\eta = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

通常 $x_0 = 1$。

不要问为何这么定义，这可以理解为 『广义线性模型』 约定的规则。

1.2. 定义二：期望估计

如果以概率论的方式解释回归（regression）这一过程，我们可以把通过给定的自变量 x，和相关的线性参数 θ 估计因变量 y 的过程。 理解为求解条件概率 $\Pr(y|x,\theta)$ 的过程。 也就是在给定了$\eta = \theta^T x$的条件下，求解因变量 y 的概率分布曲线。 然后，计算这个概率分布的期望值 $E(y|x,\theta)$，作为 y 的估计值，同时这个概率分布的方差 $Var(y|x,\theta)$作为 y 的估计值的方差。

因此第二个假设就是：y 的估计值就是 $\Pr(y|x,\theta)$ 的期望值。 如果用 $h(x,\theta)$表示 y 的估计值，这一假设写为：

$$h(x,\theta)=E(y|x,\theta)$$

1.3. 定义三：指数分布族

在毫无头绪的情况下，要求解$\Pr(y|x,\theta)$的函数表达式不太可行。因此 广义线性模型做出假设，$\Pr(y|x)$的概率分布服从如下形式：

$$\Pr(y|x,\theta)=b(y)e^{\eta^TT(y)-a(\eta)}$$

注意，这里的$\eta=\theta^Tx$，因此这个概率分布也可以写成：

$$\Pr(y|\eta)=b(y)e^{\eta^TT(y)-a(\eta)}$$

其中 $b(y)$，$T(y)$ 是 y 的函数，$a(\eta)$ 是 η 的函数。 a,b,T 这三个函数的形式未知，是一种抽象的表达方式。 一般情况下$T(y)=y$，后文中所有 $T(y)$都会直接写成y。

凡是能写成这个形式的概率分布函数，都称之为指数分布族中的一种特例。 如果 a,b,T 选择某种特殊的函数形式，指数分布族就会退化为某种特殊的概率分布（比如二项分布、正态分布），而具体的分布形式就会对应了具体的回归模型。

后文将具体介绍 a,b,T 在什么形式下会变为常见的回归模型。本文还将继续对这个GLM进行抽象地分析。

2. 广义线性模型的特征

为什么要把 y 的条件分布定义为这么奇怪的指数分布族？这是因为，在这样的定义下，我们可以证明：

- $\Pr(y|\eta)$ 的期望值

$$E(y|\eta) = \frac{d}{d\eta}a(\eta)$$

- $\Pr(y|\eta)$ 的方差

$$Var(y|\eta) = \frac{d^2}{d\eta^2}a(\eta)$$

如此简洁的期望和方差意味着：一旦待估计的 y 的概率分布写成了某种确定的指数分布族的形式（也就是给定了具体的 a,b,T），那么我们可以直接套用公式 $h(x,\theta) = E(y|x,\theta) = \frac{d}{d\eta}a(\eta)$ 构建回归模型。通过这个规律，我们可以解释为什么 logistic 回归要用sigmod 函数 $y = \frac{1}{1+e^{-z}}$ 建模。

指数分布族的这两个特征如此简洁，以至于我忍不住试着证明这两个结论。如何通过 GLM 得出常见的线性模型、logistic 回归则在后面的博客中介绍。

2.1. 证明指数分布族的期望

首先定义似然函数 $L(y,\eta)$：

$$L(y, \eta) = \log \Pr(y|\eta) = \log(b(y)e^{\eta\, y) - a(\eta)})$$

化简为:

$$L(y, \eta) = \log(b(y)) + \eta\, y - a(\eta)$$

再定义其对 η的倒数 U:

$$U(y, \eta) = \frac{d}{d\eta} L(y, \eta) = y - \frac{d}{d\eta} a(\eta)$$

可以证明 $U(y, \eta)$ 的期望 $E(U(y, \eta)) = 0$（证明过程放在后面），那么有

$$E(y - \frac{d}{d\eta} a(\eta)) = 0$$

$$E(y) = E(\frac{d}{d\eta} a(\eta))$$

因为

$$\frac{d}{d\eta} a(\eta)$$

是一个与 y 无关的函数，因此

$$\frac{d}{d\eta} a(\eta)$$

的期望就是它本身，因此

$$E(y) = \frac{d}{d\eta} a(\eta)$$

2.1.1 证明E(U) = 0

接下来证明为什么 E(U) = 0

之前定义了

$$U(y, \eta) = \frac{d}{d\eta} L(y, \eta) = \frac{d}{d\eta} \log \Pr(y|\eta)$$

带入其期望公式

$$E(U(y, \eta)) = \int U(y, \eta) \Pr(y|\eta) dy$$

中，则有：

$$E(U(y, \eta)) = \int \frac{d}{d\eta} \log \Pr(y|\eta) \Pr(y|\eta) dy$$

考虑到

$$d \log \Pr(y|\eta) = \frac{1}{\Pr(y|\eta)} d \Pr(y|\eta)$$

，有

$$E(U(y,\eta)) = \int \frac{1}{\Pr(y|\eta)} \frac{d}{d\eta} \Pr(y|\eta) \Pr(y|\eta) dy$$

化简得到

$$E(U(y,\eta)) = \int \frac{d}{d\eta} \Pr(y|\eta) dy$$

将微分符号提取出来，有

$$E(U(y,\eta)) = \frac{d}{d\eta} \int \Pr(y|\eta) dy$$

考虑到概率分布的归一化条件

$$\int \Pr(y|\eta) dy = 1$$

，则有

$$E(U(y,\eta)) = \frac{d}{d\eta} 1$$

常数的微分为0，因此

$$E(U(y,\eta)) = 0$$

2.2. 证明指数分布族的方差

先分析$E(U^2)$，带入$U(y,\eta) = y - \frac{d}{d\eta}a(\eta)$，有：

$$E(U^2(y,\eta)) = \int (y - \frac{d}{d\eta}a(\eta))^2 \Pr(y|\eta)dy$$

注意，刚刚已经证明了$E(y) = \frac{d}{d\eta}a(\eta)$，带入上式中，有

$$E(U^2(y,\eta)) = \int (y - E(y))^2 \Pr(y|\eta)dy$$

注意，y 的方差的定义就是 $Var(y|\eta) = \int (y - E(y))^2 \Pr(y|\eta)dy$

所以我们得到结论：

$$Var(y|\eta) = E(U^2(y,\eta))$$

又因为可以证明 $E(-\frac{d}{d\eta}U) = E(U^2)$（具体证明过程放在最后）。那么有：

$$Var(y|\eta) = E(-\frac{d}{d\eta}U)$$

现在要求解

$$E(-\frac{d}{d\eta}U(y,\eta))$$

，将

$$U(y,\eta) = y - \frac{d}{d\eta}a(\eta)$$

带入其中，有：

$$Var(y|\eta) = E(-\frac{d}{d\eta}U(y,\eta)) = E(-\frac{d}{d\eta}(y - \frac{d}{d\eta}a(\eta))) = E(\frac{d^2}{d\eta^2}a(\eta))$$

同理，式子

$$\frac{d^2}{d\eta^2}a(\eta)$$

与 y 无关，其期望为常数，得证

$$Var(y|\eta) = \frac{d^2}{d\eta^2}a(\eta)$$

2.2.1 证明$E(-\frac{d}{d\eta}U) = E(U^2)$

通过证明式子 $E(U^2 + \frac{d}{d\eta}U) = 0$来证明$E(-\frac{d}{d\eta}U) = E(U^2)$

之前定义了

$$U(y, \eta) = \frac{d}{d\eta}\log\Pr(y|\eta)$$

带入$E(U^2 + \frac{d}{d\eta}U)$中，有

$$E((\frac{d}{d\eta}\log\Pr(y|\eta))^2 + \frac{d}{d\eta}(\frac{d}{d\eta}\log\Pr(y|\eta)))$$

用期望公式$E(f(y)) = \int f(y)\Pr(y)dy$展开，有：

$$E(U^2 + \frac{d}{d\eta}U) = \int((\frac{d}{d\eta}\log\Pr(y|\eta))^2 + \frac{d}{d\eta}(\frac{d}{d\eta}\log\Pr(y|\eta)))\Pr(y|\eta)dy$$

因为 $\frac{d}{d\eta}\log\Pr(y|\eta) = \frac{1}{\Pr(y|\eta)}\frac{d}{d\eta}\Pr(y|\eta)$，则有

$$E(U^2 + \frac{d}{d\eta}U) = \int[\frac{1}{\Pr(y|\eta)^2}(\frac{d}{d\eta}\Pr(y|\eta))^2 + \frac{d}{d\eta}(\frac{1}{\Pr(y|\eta)}\frac{d}{d\eta}\Pr(y|\eta))]\Pr(y|\eta)dy$$

根据微分的规则：

$$\frac{d}{d\eta}(\frac{1}{\Pr(y|\eta)}\frac{d}{d\eta}\Pr(y|\eta)) = \frac{d}{d\eta}(\frac{1}{\Pr(y|\eta)})\frac{d}{d\eta}\Pr(y|\eta) + \frac{1}{\Pr(y|\eta)}\frac{d}{d\eta}(\frac{d}{d\eta}\Pr(y|\eta))$$

$$\frac{d}{d\eta}(\frac{1}{\Pr(y|\eta)}\frac{d}{d\eta}\Pr(y|\eta)) = \frac{-1}{\Pr(y|\eta)^2}[\frac{d}{d\eta}(\Pr(y|\eta)]^2 + \frac{1}{\Pr(y|\eta)}\frac{d^2}{d\eta^2}\Pr(y|\eta)$$

带入上式中，有

$$E(U^2 + \frac{d}{d\eta}U) = \int [\frac{1}{\Pr(y|\eta)^2}(\frac{d}{d\eta}\Pr(y|\eta))^2 + \frac{-1}{\Pr(y|\eta)^2}[\frac{d}{d\eta}(\Pr(y|\eta)]^2 + \frac{1}{\Pr(y|\eta)}\frac{d^2}{d\eta^2}\Pr(y|\eta)]\Pr(y|\eta)dy$$

消去相反项，有：

$$E(U^2 + \frac{d}{d\eta}U) = \int [\frac{1}{\Pr(y|\eta)}\frac{d^2}{d\eta^2}\Pr(y|\eta)]\Pr(y|\eta)dy$$

再次化简：

$$E(U^2 + \frac{d}{d\eta}U) = \int [\frac{d^2}{d\eta^2}\Pr(y|\eta)]dy$$

交换微积分顺序：

$$E(U^2 + \frac{d}{d\eta}U) = \frac{d^2}{d\eta^2}\int \Pr(y|\eta)dy$$

因为概率分布的归一化条件 $\int \Pr(y|\eta)dy = 1$，则有

$$E(U^2 - \frac{d}{d\eta}U) = \frac{d^2}{d\eta^2}1 = 0$$

得证

$$E(-\frac{d}{d\eta}U) = E(U^2)$$

参考资料：

1. http://en.wikipedia.org/wiki/Linear_regression
2. http://alumni.media.mit.edu/~tpminka/courses/36-350.2001/lectures/day32/
3. https://www.stat.tamu.edu/~suhasini/teaching613/exponential_family.pdf
4. http://www.rni.helsinki.fi/~boh/Teaching/GLMs/glmsl1.pdf

打赏